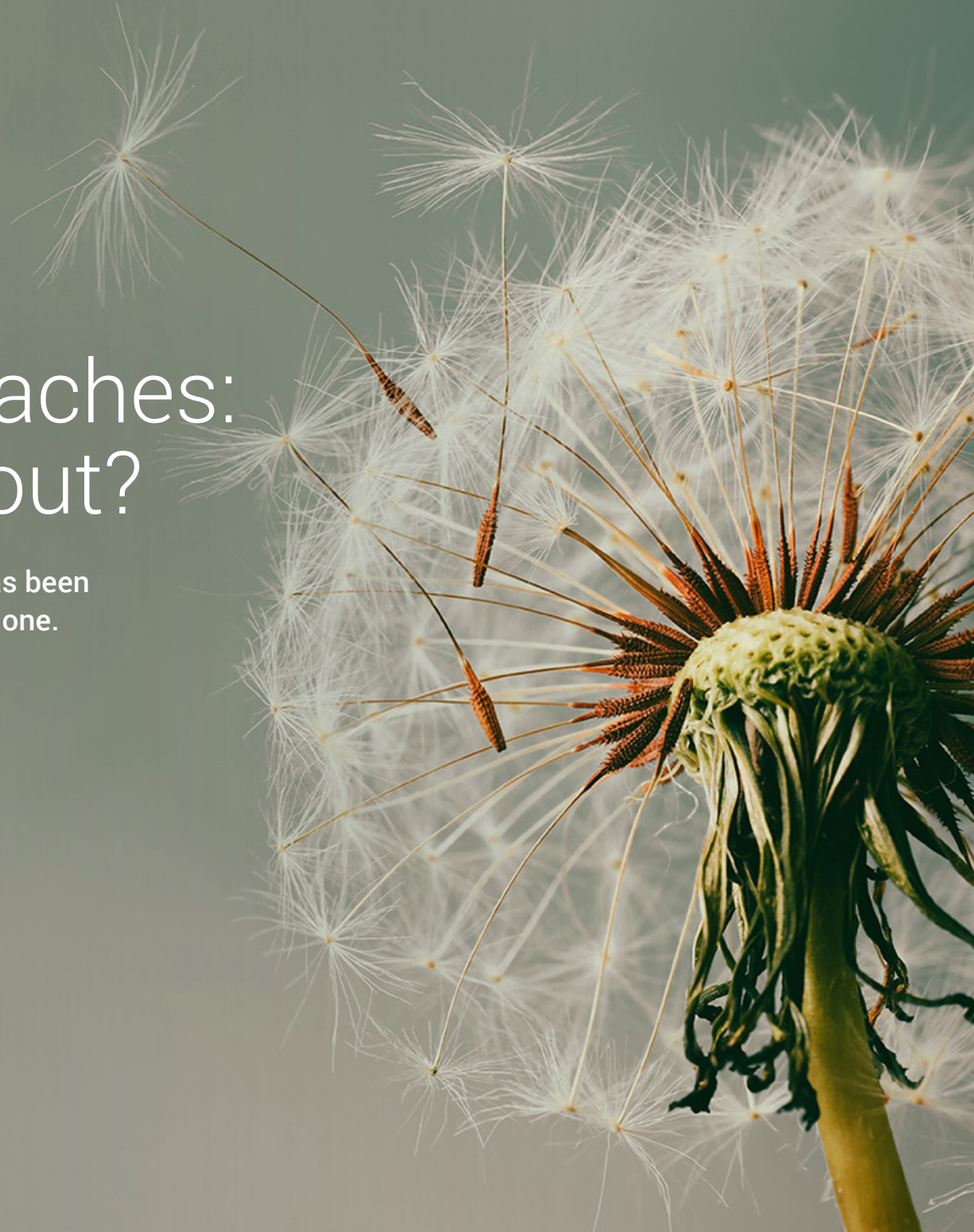


Customer data breaches: when will you find out?

Five techniques to detect when your customer data has been compromised – and respond before more damage is done.



Your data is outside your network. So how do you spot a breach?

From the moment your customer data is exposed, the clock is ticking. But with so much of your data now outside your perimeter, how would you even know?

Imagine your customers' personally identifiable information (PII) has been leaked online – and it's come from a location outside your network. Perhaps it's via a third-party supplier, or maybe an employee has shared data with a misaddressed email.

When is the first time you hear about it, and how?

This guide will show you how to spot a data breach faster – before unnecessary damage is done to your customers and your reputation, and certainly well before you receive a Google Alert or see your brand in the news.

We'll outline exactly what happens to your customer data after it's exposed and highlight the key moments when you can spot it. Then, we'll reveal five practical monitoring techniques you can use to find out sooner and respond.

Acting fast is more important than ever – and more difficult.

Compensation, reputational damage, customer attrition, fines, class action litigation... almost every cost associated with exposed PII grows by the day, until you have the problem under control. So, when a breach happens, you need to find out fast.

By the time it's on the news, the damage is already done; **just ask British Airways**. The sophisticated Magecart exploit on its booking system was live for 19 days before the story broke. In that time, the number of customers affected rose from a few thousand to an estimated 380,000. The resulting group-action compensation claim – recently settled for an undisclosed amount – is believed to be the largest in UK history, with payout estimates of up to £3 billion.



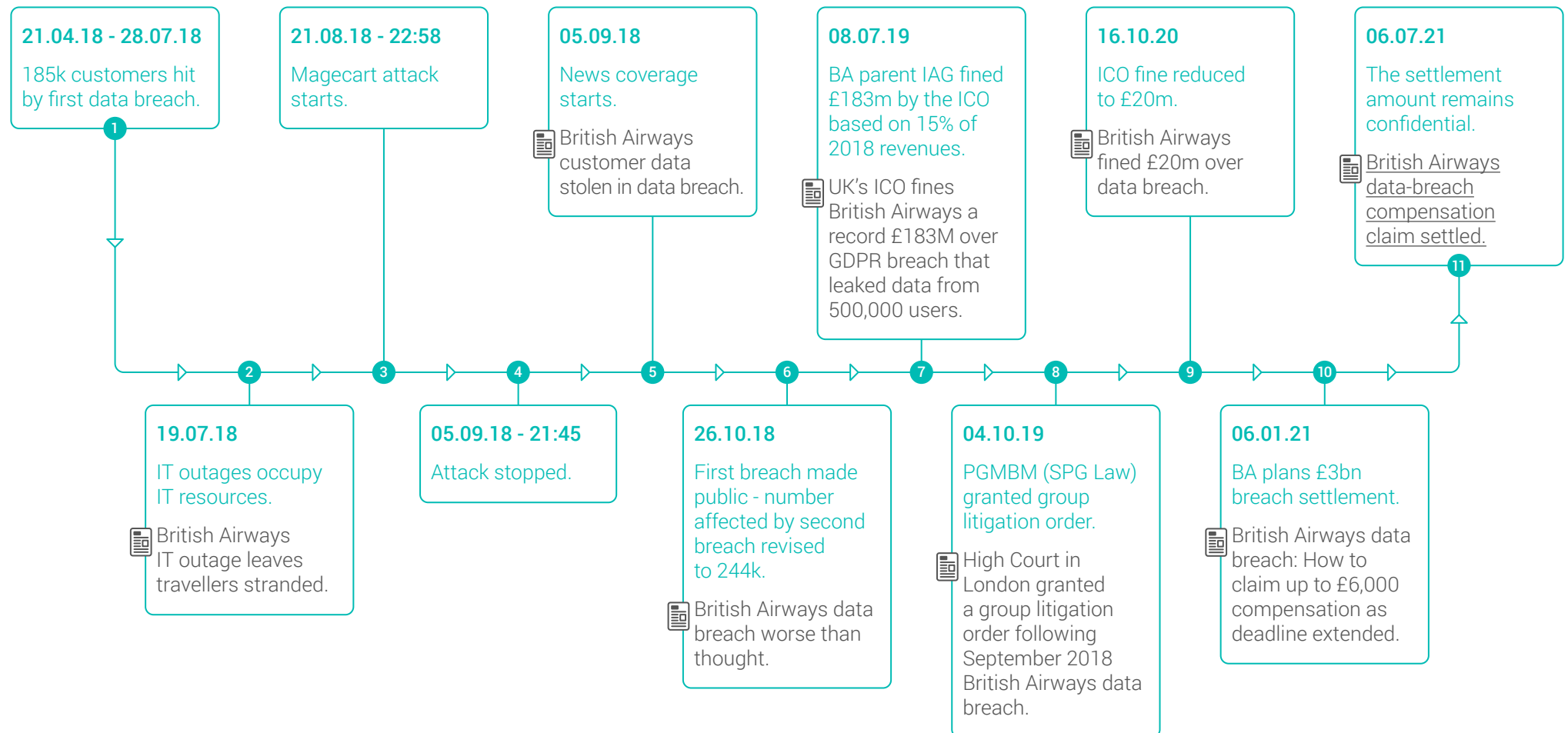
The Magecart form skimming attack that was adapted for the British Airways breach is a common exploit method. It's hard to spot – and for BA, that meant the damage kept piling up.

Magecart works by “skimming” data from your website forms, capturing login information and other data such as passport and credit card details, as the user types them in. The user's browser then sends the data to the hacker's servers, as well as submitting them to your own site as usual. So even the best firewalls and network monitoring can't detect that the data has already been siphoned off before it's reached you. The more sophisticated versions even encrypt the data and use typosquatted domains, making it even harder to spot.

The stolen data can only be detected when it's used maliciously or advertised for sale on the Dark Web or elsewhere. This shows the need for a two-pronged cybersecurity approach – simultaneously trying to prevent breaches, and monitoring for them happening.

Time is of the essence – how the £3 billion BA hack unfolded.

The financial consequences - from £183m to £20m to £3bn to confidential settlement in 11 steps.





98%

of data breaches target PII.¹

Customers expect personal data to be safe – and a fast, accurate response when it isn't.

Imagine your wallet has been stolen; that's worrying, at the best of times. Now imagine finding out it had gone missing weeks ago. **Much worse.**

The sooner you know about the problem, the better your chances of cancelling your cards before they're used extensively – and of catching fraudulent transactions on your statements. It limits the damage.

In the same way, knowing quickly when your customers' PII has been exposed enables you to reduce the impact of the breach, for them and for you.

After all, PII is not just sensitive, tightly regulated information with huge potential consequences in the event of a breach. It's also by far the most common target for hackers – accounting up to 98% of all data exposed. In 2020, the number of individual records included in data breaches more than doubled with a staggering 37 billion records exposed.²

The sooner you know, the smaller the bill.

In its 2020 *Analysis of the full costs of cyber security breaches*, Ipsos MORI lists medium and long-term costs for a cyber breach including: compensation and discounts, third-party liability, post-breach customer protection, customer attrition, and long-lasting impact on share value.

All these costs can be minimised with a faster breach response. Specifically, finding out early can enable you to:

- Limit the number of customers affected
- Prevent (or limit) follow-on incidents including phishing attacks and payment fraud, by notifying customers sooner
- Reduce damage to reputation and revenue
- Minimise business disruption and ongoing customer protection costs
- Show regulatory compliance and reduce compensation – which otherwise can be up to £18,000 per victim
- Reduce the chances of wide-scale sharing online by taking down offending posts from pastebins and forums
- Investigate the potential vulnerability in your system or supply chain, and plug it

Your data has left the building.

You can monitor your own network traffic, firewalls, and endpoints – and that's still very important. But the truth is a lot of your customer data is already out there, away from your protection and control, in:

- **Portable devices** – Whether they belong to your business or your employees, the rise in home working means there are more laptops, phones, drives, and other devices accessing your data from outside your premises.
- **Third-party apps** – On mobile devices or web-based applications in the cloud, apps have become a part of everyday productivity. Some of them you might have sanctioned, others will be "shadow IT" your users have sourced for themselves.
- **Cloud services and SaaS** – Even for the most secure business, IT is almost unimaginable now without at least some reliance on cloud storage and SaaS systems.
- **Your supply chain** – When you share your customers' data with a supplier – and they share it with theirs in turn – it remains your responsibility. As the 2020 SolarWinds Orion attack shows, even the best vetting is no guarantee of security.

Simply, the nature of modern IT means that – for an organisation of any size – your customers' PII is never confined to any one system, network, or premises. For most, it's likely stored and processed across hundreds or even thousands of systems.

Wherever that data lives, you're responsible for it. And, unless you monitor independently, you're relying on any one part of the chain to notice – and to inform you promptly – when a breach has taken place.

"We store data about our clients in around a thousand different systems."

CISO, Travel Industry, UK

Why traditional cybersecurity systems don't spot breaches.

There are now two approaches to protecting your customer data. One is preventing a breach – protecting your perimeter, auditing your supply chain, and training your people. The other is finding out, quickly, when something has gone wrong and your data has been exposed.

The problem is, how would you know – especially now so much of that data is stored and processed out in the world, far beyond what you can see on your network?



Protect the data where you can – and monitor for it where you can't.

You're already working hard to protect your customers' data, and rightly so. But you can't prevent every breach – especially where you have an extended supply chain – so it's also important to identify your data quickly when it's exposed.

To do that, you need a separate set of tactics to supplement your usual cybersecurity plans. And that's what the rest of this guide is about.

First, we'll focus on what happens to customer data after a breach: how and when it changes hands on the Dark Web, and the key moments you can spot it early and intervene to prevent any further damage.

Then, we'll share five practical monitoring techniques which you can use to identify your customer data and determine where the leak came from. Some are very simple to perform manually, or at low or no cost. Others require a more sophisticated tool set – but even there, it doesn't have to be complicated.



“Hackers don’t break in; they log in.” Bret Arsenault, CISO, Microsoft³

Humans are still human – and their mistakes are hard to detect.

Meanwhile, human beings remain accident prone. People still love saving data in spreadsheets and carrying it outside your walls. They still leave laptops on buses, and portable drives in cafés. Sending an email to the wrong address remains one of the most commonly reported kinds of breach. It’s no surprise that, according to the ICO,⁴ human error accounts for 75% of all data security incidents.

75%
of data security incidents are
attributable to human error.⁴

Other common mistakes include leaving ports open or files unprotected on servers that can be accessed without a password. And insider threats are a growing issue: alongside the established risk of disgruntled current and former employees, Cifas has detected a spike in “Fraud-as-a-Service” schemes⁵ among financial services staff.

None of these issues fit the conventional idea of an external party using a technical exploit to breach your defences - so they’re difficult for traditional cybersecurity measures to spot. How do you track data that’s been emailed to the wrong address?

Stolen credentials don’t trigger alerts.

When malicious actors do target your organisation, they’re likely to use increasingly sophisticated spearphishing, typosquatting, and social engineering techniques to obtain genuine login credentials.

Once they have the keys, they simply access your data through the front door. Microsoft CISO Bret Arsenault was right: hackers don’t break in anymore – they log in.

One key problem is that this approach is unlikely to trigger an IT, network or cybersecurity alert. So, you’re unlikely to ever know.

What happens to your customer PII after it's breached?

As soon as your customer data gets into the wrong hands, a process has begun.

And the longer it takes you to find out about it, the greater the damage – to your customer, and to your organisation.

Although the exact timing and details may vary, breached data often goes through a process much like this:

During or immediately after the attack

The hacker either attempts to use the data themselves – typically for fraud, phishing, or spam – or advertises it for private sale through end-to-end encrypted channels. These transactions are extremely difficult to detect without specialist monitoring.

Week 1

The data may be promoted to a broader audience. It could be advertised on Dark Web paste-and-dump sites, or mentioned on hacker forums and chatrooms. Meanwhile, the original hackers or buyers may be matching the data with other breaches to package up more lucrative, complete identity kits.

Month 1

Hackers start trading the data among themselves. It is likely listed on paste sites and forums and bartered in exchange for other data sets. Advertisements might be posted online – for example on Reddit or Telegram.

Month 3

The data is passed around and repurposed – for example to see if one set of email and password credentials can access multiple services or accounts. It is compiled into larger, combined lists on mainstream forums and used for credential stuffing.

Month 6+

At some point the data itself – or a sample of it – is shared online without cost, providing access to the infosecurity community. News of the data breach is either leaked intentionally to damage the company's reputation, or its presence is suspected by customers who have been targeted, and it becomes a major news story.

Month 12+

Even after it's discovered, a significant breach – with a large number of records and plaintext passwords – can circulate online for months or even years. The internet has several well-known “classic” breaches, and parts of them are routinely added into newer datasets to pad out the number of records. Because of this practice, exposed records can be reused almost indefinitely.

When are the best opportunities to spot a data breach?

The typical journey after a breach presents four key moments when you have the best chance to spot your data, trace its source, and take action:

1. When the data is misused.

Whenever a hacker or buyer uses the stolen data, there's an opportunity to spot the leak. For example: the online bank Monzo was the first to spot the 2018 TicketMaster breach,⁶ when it noticed that many card holders reporting fraudulent transactions had also bought tickets.

2. When it's advertised on the Dark Web.

Monitoring the sites and forums where data is offered for sale gives a chance to intervene before a dataset spreads too far. Usually, this is achieved by working with analysts and researchers, or using an automated Digital Risk Protection platform which can give greater coverage.

3. When it's discussed on the open web.

Free and low-cost tools like HaveIBeenPwned⁷ give an inexpensive way to know if an email address or phone number has been included in a breached data set. This is useful but, because they focus on publicly available data, can reveal the breach relatively late on.

4. When it becomes a news story.

Google Alerts and social listening tools can tell you when the media is discussing your brand in relation to a breach. By this time, most of the damage has already been done.

Clearly, techniques which enable you to spot breached data earlier in the process are most helpful for limiting costs to your revenue and reputation, as well as disruption for your customers. However, in practice each breach can follow a slightly different route – so it's important to take a combined approach, covering as many bases as you can.

Five ways to spot a customer data breach early.



Happily, there are a number of practical techniques you can take to discover when your customer data has been exposed – before it becomes a news story. And many of them are quick and easy to implement.

As well as spotting a potential leak, there are three important considerations which you'll need to take into account with the combination of methods you choose:

- **Proving whether it's your data**

It's important to know for certain whether the exposed records originated from your company or if, for example, it's a different source that happens to have some of the same customers.

- **Identifying the time and location of the leak**

Especially if you have a wide supply chain or a complex IT estate, you need to be able to trace exactly where, when, and how the data was breached so you can take the right action.

- **Ensuring your methods are GDPR compliant**

The process of searching for exposed customer PII must itself meet the standards for data use. You need to take care not to increase the chance that the data will be exposed.

The five techniques we will share here can work together to help you meet these requirements.

Some of them can be performed yourself, manually, or by working with data security colleagues. Others need a greater level of automation and scale – for example, using a Digital Risk Protection (DRP) platform that's designed for the purpose.

What is Digital Risk Protection?

A DRP platform gives you an automated way to scan the surface, deep, and Dark Web for your business data, as well as potential threats.

It lets you take a more holistic approach to your cybersecurity by monitoring the landscape outside your network defences. That means you can:

- Spot suspicious activity like typosquatting and Dark Web chatter
- Focus your defences to meet specific, real-world threats
- React faster to data leaks – from any source – and limit further damage

As such, it's an ideal way to deliver some of these techniques easily, and at scale, while minimising extra workload for your team.

Five ways to spot a customer data breach early.

Technique #1

Add synthetic “BreachMarker” entries to watermark your datasets.

When it works: Whenever the breached data is misused.

This is a simple but effective technique. You seed each dataset with a unique, fictional record, which acts as a marker. Any unexpected activity involving this ghost customer gives you an instant, definitive sign that your data has been leaked.

We call this record a “BreachMarker”, and it’s quick and easy to deploy.

The approach has several key advantages. Most importantly, it’s potentially extremely fast – you can detect that the data has been misused even if it has never been advertised for sale. All you need to do is monitor for unexpected email.

It also eliminates “false positives”. Because the identity you insert is synthetic, it has no accounts with other companies and there are no other possible sources for the data.

Helpfully, the fact there is no real person involved means the GDPR does not apply when you detect the record’s misuse.

But like any statistical approach, it can’t always be relied on in isolation. And for some datasets it may be difficult to insert a synthetic identity – for example if “know your client” checks are required.

To take things a step further, you can rotate different combinations of BreachMarkers over time, and include different ones for each supply chain partner. You can then compare this with any exposed data, to pinpoint the exact source and timing of the breach.

Five ways to spot a customer data breach early.

Technique #2

Monitor the Dark Web for your company's brand name.

When it works: When the data is advertised for sale.

When hackers offer data for sale, they will often mention the name of the company where it originated. This gives you an opportunity to step in.

Whether manually or using an automated platform, you can monitor for your brand name on the Dark Web. Remember to include multiple variants, sub-brands and misspellings to make sure you catch all the relevant conversations.

Hacker forums are also a good place to start – though ransomware increasingly exfiltrates data as well as locking it, as an added threat to data owners. So, it's a good idea to check ransomware sites too.

“Notable in 2020 was the growing trend of ransomware operators threatening to leak data from victim organisations, and in some cases actively doing so.”

CrowdStrike, Global Threat Report 2021⁸

While this approach can help you to spot a breach relatively early, it is also prone to false positives. Contributors will advertise data they don't actually have, misrepresent the source of the records or simply be in the business of getting you to click on malicious links.

It's therefore important to be able to filter the search results and discover quickly whether a post is both relevant and genuine. The credibility of the seller is another key factor; understanding the profile and history of a post's author can help you to gauge whether they're likely to have the data they claim.

In many cases – particularly where a ransomware gang is threatening to leak customer PII – a sample will be included. By analysing this, you can verify whether the data could indeed be yours.

Five ways to spot a customer data breach early.

Technique #3

Check uploaded data for strings specific to your database.

When it works: When data is uploaded to dump sites.

The PII content of a leak is not the only – or even necessarily the best – way to confirm where the data came from.

Certain character strings are unique to your organisation's database. And these can often be included when data is dumped online or exported as a CSV, or if source code is accidentally exposed on a code repository.

So, alongside brand and customer data details, you can search for things like:

- Database names
- Table names
- Server addresses
- IP addresses

Likewise, you can look for patterns and formats that are unique to your data – for example, if your customer ID codes or reference numbers contain a certain combination of numbers and digits, or they're arranged in a specific or unusual way.

This technique has two key benefits. First, a string or format that's unique to your database infrastructure is unequivocally yours – unlike a customer email address or password, which they might use on several sites.

And second, searching for these strings is safe to do. Because you're looking for metadata, rather than the data itself, there are no GDPR issues and you don't need to expose any PII as part of your search.

Five ways to spot a customer data breach early.

Technique #4

“Fingerprint” your data by comparing with exposed datasets.

When it works: When data is leaked or uploaded to dump sites.

This is a non-invasive way to monitor for a breach without inserting any additional records into your data. It involves continuously comparing exposed datasets against your entire database, to discover how much they overlap.

If none or only a few of the records match your customer base, the dump file is generally safe to ignore. However, if there's a significant overlap – or if the entire dump is a subset of your customer records – it definitely needs to be investigated further.

To screen out false positives, it's sensible to filter results by both the total number of email addresses in common, and the percentage of the dump which are also your customers.

Although it's a simple idea in theory, it requires scale, especially if you have a large customer base with millions of records. And if these records are active and updated regularly, you'll definitely need automation.

But using an automated platform needs to be secure, so you can check against your entire customer database without risking that it will be breached. For this reason, customer details should undergo one-way encryption, or hashing, so the platform can make the comparison without storing the original data.

However, hashing is not secure in isolation. It's prone to dictionary attacks, based on hashing large tables of common login credentials. Free websites are available for this specific purpose, and it's an easy way to crack common passwords.

We therefore recommend you add another layer of complexity by “salting” the data with a secret, additional string of characters before it's encrypted – so cracking the hash requires the algorithm used, the original data, and the secret character string.

This means any data uploaded to the platform is always protected.

Five ways to spot a customer data breach early.

Technique #5

Take fingerprinting further by searching for anonymised data.

When it works: When data is leaked or uploaded to dump sites.

To further enhance security, this approach uses a “fuzzy search” for a fragment of fingerprint data, instead of aiming for a 1:1 match with a complete encrypted record. As a result, any data used outside your system is always completely anonymised.

The approach intentionally returns more false positives, so it’s important to adjust the filter thresholds to understand which results merit a closer look.

For maximum security, you can carry out all encryption and anonymisation steps of the process within your own trusted environment. Then take excerpts and discard the rest, to produce fingerprint fragments, which you can upload to your DRP platform using a “Write Only” API.

Although it sounds complicated, this can be as simple as updating a spreadsheet, and exporting the anonymised data. To protect your ‘live’ datasets, you can take it a step further, integrating REST APIs to completely automate the update process.

The platform’s automation can then monitor for these fragments across the surface, deep, and Dark Web as usual. Data matching the “fuzzy search” is brought back into your secure environment, where any false positives are discarded automatically. Genuine results matching your filter criteria then generate alert notifications, letting you safely analyse the incident in the security of your own environment, without ever exposing any PII to a third party.

For the best results, use a combination of techniques.



These methods are all practical and highly effective in their own right – but they're best used in conjunction with each other, to give the widest coverage and the best chance of intercepting breached data as it spreads.

Using them together can also give you a clearer picture of what's happening. For example, you might use Fingerprinting to identify data dumps which could be of concern, and then search those files for BreachMarkers to confirm the source of the data.

If you can use all five techniques together, you'll be well placed to respond far earlier when a breach occurs and protect your customers and your business from further harm.

Time for a more realistic approach to protecting customer data.

Whichever methods you choose, the most important thing to adopt is a new mindset when it comes to protecting your customers' PII.

As the number and sophistication of attacks – and the length of IT supply chains – continue to grow, customer data breaches will only become more likely, and more difficult to detect.

Although cyber defences remain important, the truth is that the traditional idea of the network perimeter is fast becoming obsolete. Your data is everywhere, so protecting it takes a more holistic approach. You need to look outside your organisation to detect threats and compromised data.


Basic, manual measures like some of those outlined in this guide are a good start. But the Dark Web is huge, and this task is ultimately best performed at scale, with the help of tools designed for the job.

Skurio's Digital Risk Protection platform has options for companies of all sizes – with built-in security and GDPR compliance, and smart automation that takes the headache out of proactive monitoring.

Importantly, our experts are always here for a straightforward, no-fuss conversation to help you work out what measures you really need. If you'd like to explore the subject further – or if you have questions about any of the techniques in this guide – feel free to call us for an informal chat.

Call us today on **+44 28 9082 6226**
or email **info@skurio.com**

www.skurio.com



SKURIO LTD | ARTHUR HOUSE | 41 ARTHUR STREET | BELFAST | BT1 4GB
+44 28 9082 6226 info@skurio.com skurio.com